

Data Management System for Ab-Initio Nuclear Physics Calculations - MFDn

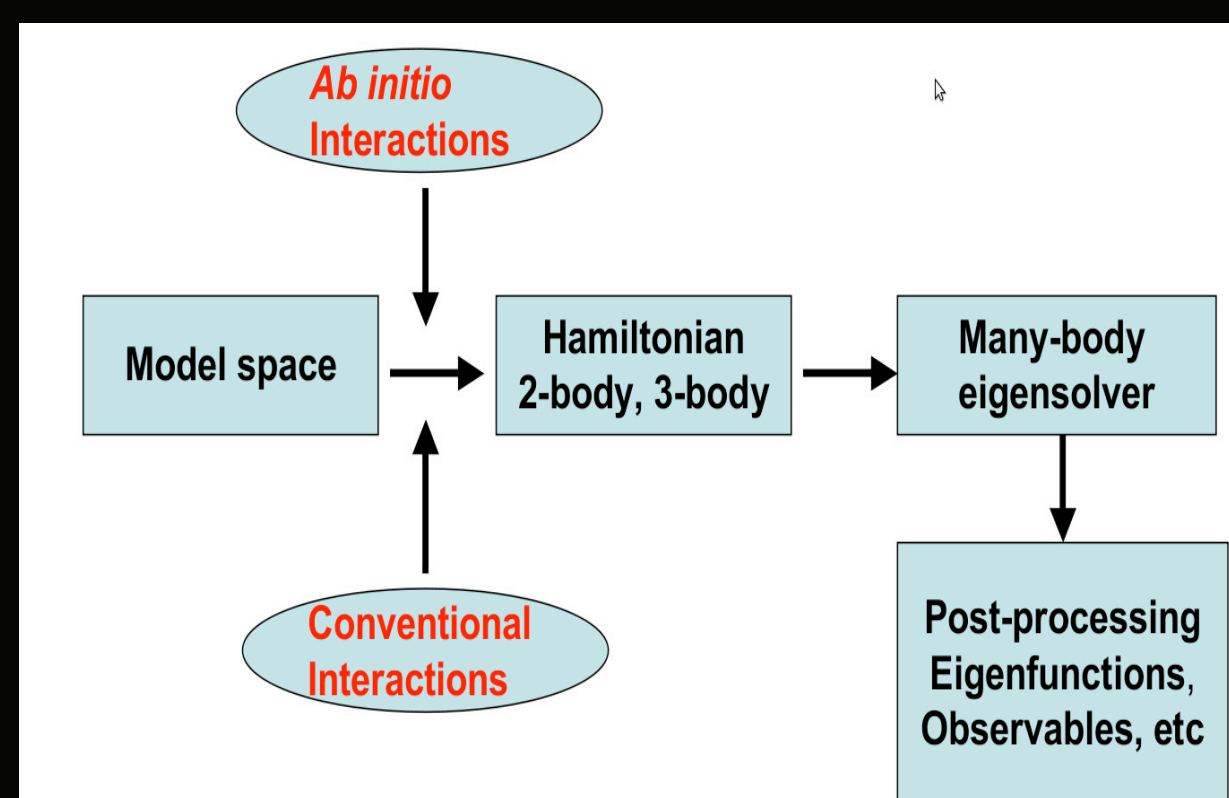
MFdn Background

Many Fermion Dynamics for nuclear structure (MFDn) software package is developed by James Vary's group at Iowa State University.

In MFDn, the nuclear Hamiltonian is evaluated in a large harmonic oscillator basis and diagonalized by iterative techniques to obtain the low-lying eigenvalues and eigenvectors.

Functionalities:

- Generate Many-Body (MB) states from single-particle states.
- Construct Many-Body Hamiltonian from NN and 3N interactions, it generates large sparse unstructured matrix.
- Diagonalize Hamiltonian.
- Optional: Calculate physics observables



Motivation

Large volume of data is generated per experimental run, and rerunning the experiment may take a long time.

For validation purpose, later version of code may need to compare the output with the early version's results.

Data analysis may be done by the people who do not run the experiment.

Challenges and requirements

A data management system needs to be built to store the provenance information to facilitate the reproducible research, but few challenges to be addressed are as follows:

- How to record each run, tightly coupled or loosely coupled approach?
- How to store the large volume of data in database?
- How to group the run information for easy retrieval?
- How to reference the partially processed information?

Data Search over WWW

Information on Existing Nuclear Calculations
[Nuclear Physics Server Home | DBMS Home | DBMS Search]

Enter search criteria:

runID

username

machineID

2B_potential

3B_potential

4B_potential

ext_field

Z

N

nshell_min_Z

nshell_max_Z

nshell_min_N

nshell_max_N

Nmax

parity

twiceMj

twiceJ

Nstates

refstate

J

T

Info_file

This form is input by users from online webpages, the fields include the physics properties for search criteria. The below form is the result page from the search.

Information on Existing Nuclear Calculations
[Nuclear Physics Server Home | DBMS Home | DBMS Search]

385 records found.

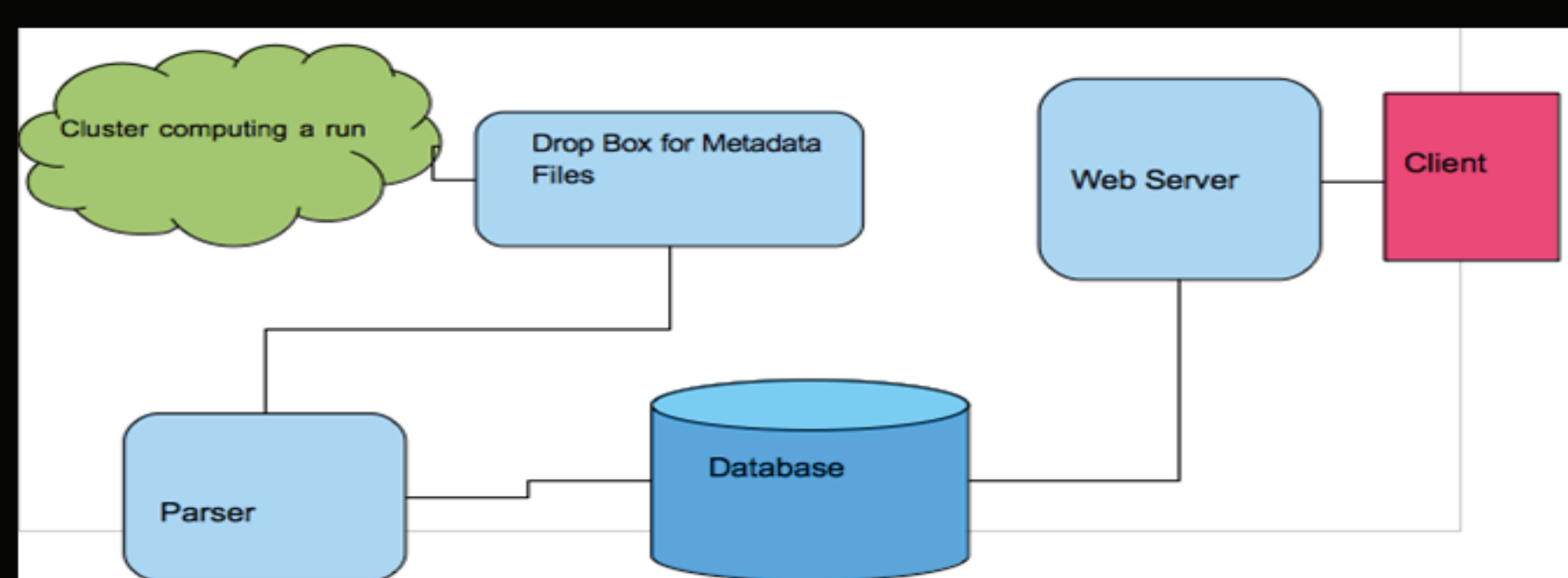
runID	info_file	username	Z	N	nshell_min_Z	nshell_max_Z	nshell_min_N	nshell_max_N	Nmax	Nstates	twiceJ	twiceMj	START_DATE	END_DATE
297	mfdn.info.d100413103308	pmaris	3	4	1	6	1	6	4	10	-1	1	2010-04-13 10:33:08	2010-04-13 10:34:10
298	mfdn.info.d100414081310	jvary	2	2	1	11	1	11	10	10	-1	0	2010-04-14 08:13:10	2010-04-14 09:26:19
299	mfdn.info.d100414093910	jvary	2	2	1	11	1	11	10	10	-1	0	2010-04-14 09:39:10	2010-04-14 09:39:10
300	mfdn.info.d100420105605	jvary	2	2	1	11	1	11	10	10	-1	0	2010-04-20 10:56:05	2010-04-20 12:15:01
301	mfdn.info.d100420193428	jvary	2	2	1	11	1	11	10	10	-1	0	2010-04-20 07:34:28	2010-04-20 08:52:23
302	mfdn.info.d100421223615	jvary	2	2	1	11	1	11	10	10	-1	0	2010-04-21 10:36:15	2010-04-21 11:54:04

Supporting HDF5

Highly efficient output format for MFDn is HDF5. This is made compatible with DBMS. It also provides useful information about wavefunction like sparsity structure and dimensions of vector. This information is uploaded in the database by converting HDF5 headers into XML format and then parsing them using simpleXML.

Data Management System Architecture

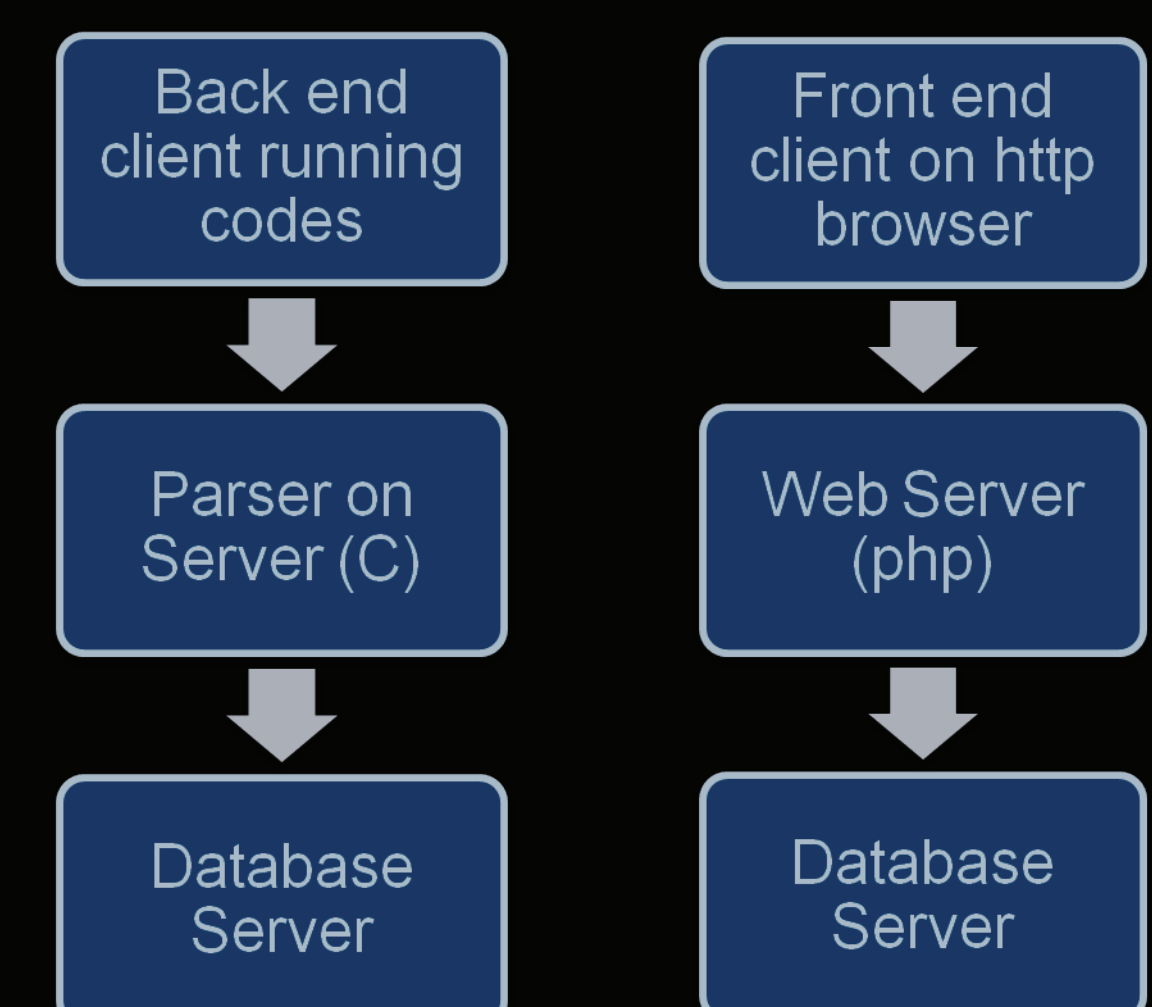
- Drop Box: an info file is output of each run of MFDn which is placed in a hot folder on dbms server. An info file consists of all data about the run, e.g. compilers, version of code, machine name, kernel version, number of processors used, other information includes identifying characteristics of the run, e.g. number of nucleons, Nmax, results about the run.
- Parser: parses the mfdn.info file and inserts the run record to database; serves as the backend for web-based retrieve functions.
- Web Server: web based front end searches and lists the existing run, and allows the update on file paths.
- Database server: stores all the related metadata for each run of MFDn.



System Running Workflow

Left: A daemon runs periodically to parse mfdn.info file and creates instance of structure 'run', connects to sql server and hits sql query to upload contents of this instance.

Right: Retrieve the record from database and display it on webpage. Http requests are passed to web server through search form, a search result is returned.



Conclusions and Future work

A data management system is built for Ab-initio Nuclear Physics code - MFDn. The experimental data and suitable provenance information such as scripts, compilers, hardware are included in the database, so that an expert researcher could reproduce the data. HDF5 files are also supported as well as binary output format. In the future, we would like to apply this system to another Ab-initio Nuclear Physics code - Bigstick. A common data management system API will be proposed to the community to gain consensus.

Reference

"Accelerating Configuration Interaction Calculations for Nuclear Structure", Philip Sternberg etc, SC 08 proceedings.